

## REMARKS

Claims 1-3, 5-7, 9-10, and 12-18 were pending in this application. In this response, the Applicant has amended claims 1-2, 6, cancelled claims 3-5, 7-14, 16-21, and added claims 22-34. Accordingly, claims 1-2, 6, 15, and 22-34 remain pending.

A Preliminary Amendment was filed on April 2, 2008 which amended original claim 1 and canceled claims 2-21. The Office Action mailed April 7, 2008 rejected claims 1-21 as the Preliminary Amendment was not entered. In response to the interview with the Examiner conducted on December 24, 2008 claim 1 is herein amended to incorporate additional limitations similar to those previously presented in the Preliminary Amendment which was not entered.

The Applicant respectfully submits that the present application is in condition for allowance.

### *Claim Rejections*

The Office Action rejected claims 1-3, 5-7, 9, 10 and 12-18 under 35 U.S.C. 103(a) as being as being unpatentable over U.S. Patent Application Publication Number 2004/0030741 of Wolton et al. (hereinafter "Wolton"), in view of U.S. Patent Application Publication Nurnber 2004/0064471 of Brown et al. (hereinafter "Brown"), and in further view of U.S. Patent Application Publication Number 2005/0060295 of Gould et al. (hereinafter "Gould").

Claim 1 has been amended to recite (emphasis added):

1. (currently amended) A computer-implemented method for information retrieval, classification, indexing, and summarization, comprising:  
"identifying a collection of hyperlinked documents as a single coherent compound document on a single topic created by a number of collaborating authors, wherein the identifying includes observing results of a first number of heuristics run on the collection of hyperlinked documents and related hyperlinks, wherein the first number of heuristics includes identifying at least one of: similar creation dates and similar last-modified dates; and wherein the collection of hyperlinked documents is distributed over a plurality of URLs, wherein the first number of heuristics includes:

identifying hyperlinks that link within a same directory and include a sufficient quantity of common anchor text,  
identifying hyperlinks that contain linguistic structures that indicate relationships between document parts,  
identifying external hyperlinks to same places,  
identifying at least one of: similar creation dates and similar last-modified dates,  
identifying individual URLs having similar structure indicating an order of inclusion in the compound document, and  
identifying a link structure of "wheel" form;  
analyzing the content and structure of the compound document to find a preferred entry point for the compound document, wherein the analyzing includes observing results of a second number of heuristics run on the compound document and related hyperlinks, wherein the analyzing includes and combining the results of the second number of heuristics run on various hyperlinked documents of the compound document, wherein the results of the second number of heuristics include numerical scores, wherein and the combining includes a weighted averaging of the numerical scores into an overall score, and wherein a maximum overall score determines the preferred entry point; and wherein the second number of heuristics includes:  
identifying specific filenames that define entry points, including at least one of: "index" and "default".  
identifying a particular component document in the compound document as a suitable entry point because the component document has several in-links, wherein the in-links are from outside the compound document.  
determining a measure of vector distances along intra-document links between a particular component document and all other component documents in the compound document.  
determining whether a URL has links pointing to longer URLs having common directory components followed by different ending directory components, wherein the ending directory components contain specific identifying information;  
processing the compound document as a whole, wherein processing the compound document as a whole includes including at least one of: indexing, classification, and retrieval; and  
processing the compound document from the entry point, wherein processing the compound document from the entry point includes including at least one of creating at least one of: a presentation of results from retrieval, summarization, and classification, the entry point as a representative URL for the compound document.

The Applicant respectfully submits that claim 1, considered as a whole, is patentable over Wolton, Brown and Gould.

For example, claim 1 requires, *inter alia*, "the collection of hyperlinked documents is distributed over a plurality of URLs, wherein the first number of heuristics includes: identifying hyperlinks that link within a same directory and include a sufficient quantity of common anchor text, identifying hyperlinks that contain linguistic structures that indicate relationships between document parts, identifying external hyperlinks to same places, identifying at least one of: similar creation dates and similar last-modified dates, identifying individual URLs having similar structure indicating an order of inclusion in the compound document, and identifying a link structure of "wheel" form," and "wherein the second number of heuristics includes: identifying specific filenames that define entry points, including at least one of: "index" and "default", identifying a particular component document in the compound document as a suitable entry point because the component document has several in-links, wherein the in-links are from outside the compound document, determining a measure of vector distances along intra-document links between a particular component document and all other component documents in the compound document, determining whether a URL has links pointing to longer URLs having common directory components followed by different ending directory components, wherein the ending directory components contain specific identifying information." Wolton, Brown and Gould do not specifically disclose these limitations.

Wolton does not describe these limitations of claim 1. Wolton describes "A modular intelligent personal agent system...for search, navigation, control, retrieval, analysis, and results reporting on networks and databases." (Wolton, Abstract.) Wolton describes that what distinguishes it "from other search and retrieval agent systems available for application to the World Wide Web, is that it provides a open ended flexible agent creation and configuration tool that does not require any programming experience

to use, and thereby permits non-programmer users the ability to generate sophisticated web search and retrieval agents and suites of agents." (Wolton, ¶0169.)

Brown also does not describe these limitations of claim 1. Brown describes "A method for presenting content from the page in a distributed database." (Brown, Abstract.) In Brown's a preferred embodiment, "a server receives a request from a client for a page from the database wherein the page has a plurality of links to linked pages in the database. The server retrieves the page and generates a set of thumbnails of the linked pages in the database. The server then sends the page and the set of thumbnails to the client." (Brown, Abstract.)

Gould also does not describe these limitations of claim 1. Gould "relates to network communication systems, and more particularly to statistical classification of network data for signature-based security and quality-of-service." (Gould, ¶0002.) Gould describes "A network data classifier [that] statistically classifies received data at wire-speed by examining, in part, the payloads of packets in which such data are disposed and without having a priori knowledge of the classification of the data." (Gould, Abstract.)

Wolton, Brown and Gould do not disclose the limitations in combination. Even if it would have been obvious to one of ordinary skill in the art at the time of the invention to combine Wolton with Brown and Gould, Wolton, Brown and Gould still fail to render obvious the limitations because Wolton, Brown, and Gould do not disclosure these limitations of claim 1 in combination.

Accordingly, the Applicant respectfully submits that claim 1, considered as a whole, is patentable over Wolton, Brown, and Gould. Claims 2-3, 5-7, 9-10, 12-18, and 22-34 each depend directly or indirectly from claim 1. Therefore, claims 2-3, 5-7, 9-10, 12-18, and 22-34 are patentable over Wolton, Brown and Gould for at least similar reasons.

Claims 2, 6, 15, and 22-34 each depend directly or indirectly from claim 1. Therefore, claims 2, 6, 15, and 22-34 are patentable over Wolton, Brown and Gould for at least similar reasons. Claims 3-5, 7-14, 16-21 are canceled in this response.

Accordingly, the Applicant respectfully requests withdrawal of the rejections of claims 1-2, 6 and 15 under 35 USC 103(a).

Furthermore, new dependent claims 25, 26, 29, and 32 require that “processing the data set of URLs further includes resolving redirects within a directory,” or a similar limitation. The Applicant respectfully submits that claims 25, 26, 29 and 32, considered as a whole, dependent on claim 1, is patentable over Wolton, Brown and Gould. Wolton, Brown and Gould do not disclose this limitation. Wolton teaches a method for a user in conjunction with a graphical user interface to initiate “redirects” to reach the origin server of a web page addressed and served via a secondary server” (Wolton, ¶¶[0705]). Wolton does not resolve URL redirects within a directory but rather enables redirects. Furthermore, Brown and Gould do not disclose the limitation of “processing the data set of URLs further includes resolving redirects within a directory.”

New dependent claims 28 and 32, as dependent on claim 1 requires “removing an argument part of a URL” which “follows a # symbol or a ? symbol.” The Applicant respectfully submits that claim 28 and 32, considered as a whole, dependent on claim 1, is patentable over Wolton, Brown and Gould. Wolton, Brown and Gould do not disclose this limitation. Wolton discusses using a ? symbol as part of a navigation button (icon link) not as a part of a URL or to remove parts of a URL that follow a ? symbol (Wolton, ¶¶[0333], Fig. 4, 804; ¶¶[0411]; ¶¶[0697]). Wolton discusses using a # symbol as particular mark that represents a numeral convention, not as a part of a URL or to remove parts of a URL that follow a # symbol (Wolton, ¶¶[0180], ¶¶[0341], ¶¶[0342], ¶¶[0348], ¶¶[0777], ¶¶[0837], ¶¶[0838]) Thus, Wolton does not disclose the limitation “removing an argument part of a URL” which “follows a # symbol or a ? symbol.” Furthermore, Brown and Gould does not disclose the limitation “removing an argument part of a URL” which “follows a # symbol or a ? symbol.”

**CONCLUSION**

The Applicant respectfully submits that the present application is in condition for allowance.

In the event a telephone conversation would expedite the prosecution of this application, the Examiner is invited to call the undersigned at (408) 927-3380 Although no fee is believed to be due, the Commissioner is authorized to charge any such fees in connection with the filing of this paper to Deposit Account No. 09-0441 (Order No. ARC920030028US1).

Respectfully submitted,

Date: January 9, 2009

By:



Mohammed Kashef

Reg. No. 60,762

Intellectual Property Law  
IBM Almaden Research Center  
650 Harry Road  
San Jose, California 95120  
Telephone: (408) 927-3380  
Facsimile: (408) 927-3375  
Kashef@us.ibm.com